Análisis de contenido Audiovisual

Caso de estudio - streaming C5N y TN

Trabajo Final DataScience - Grupo 2 - DH Lima - 2020

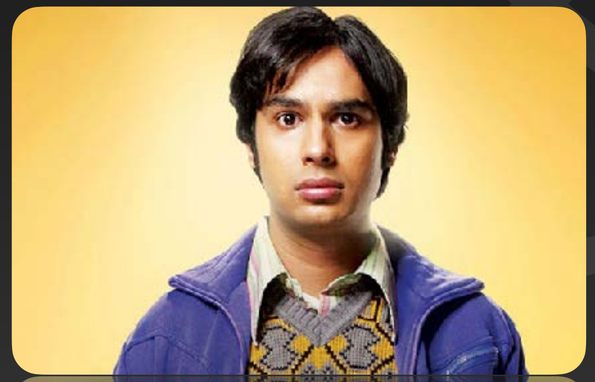
DigitalHouse > Coding School

Integrantes

The Big Band Theory

Franco Pelecchia

Diego Lesertesseur



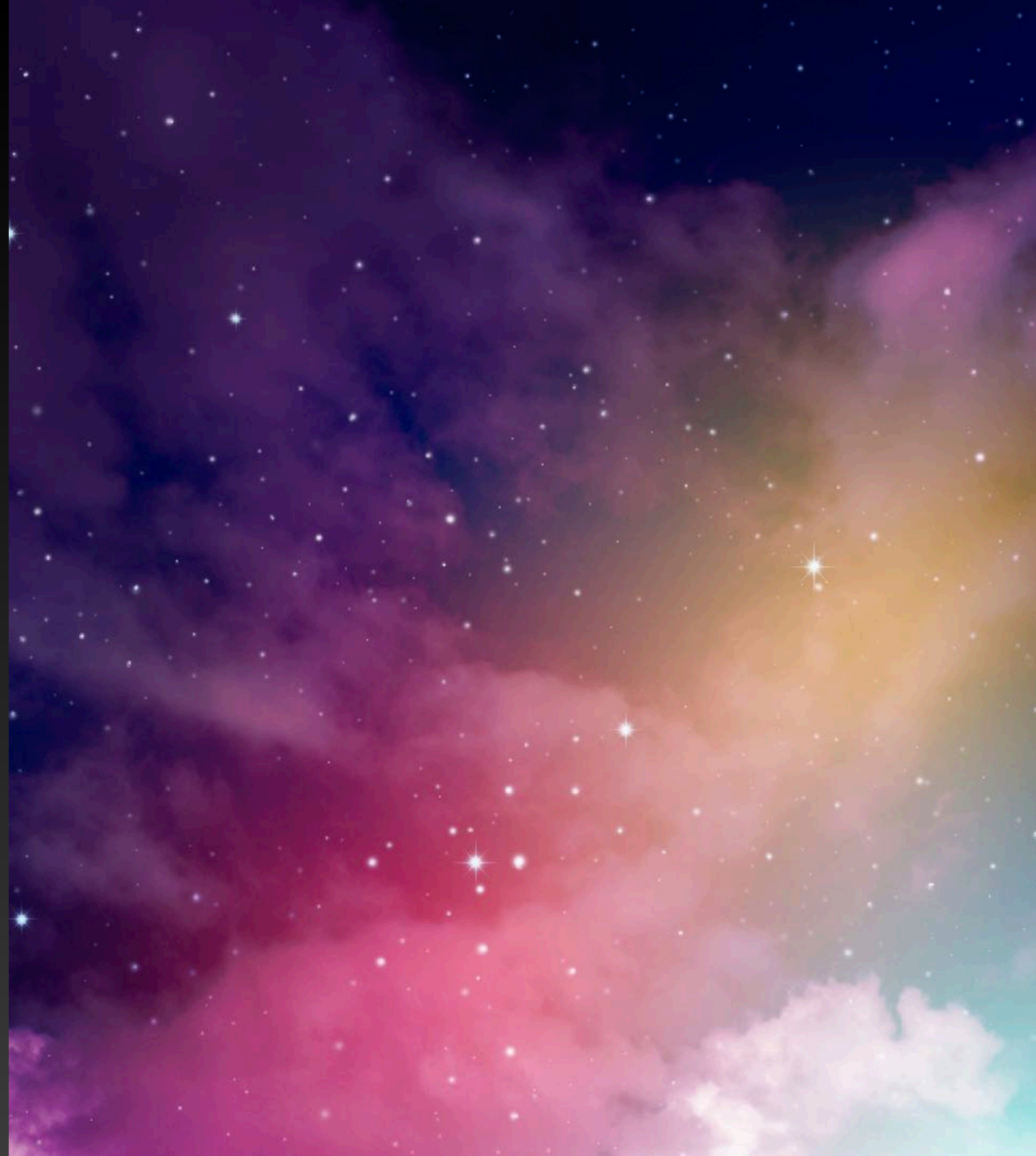
Horacio Suarez

Néstor Magariños

“La teníamos a Camila Panizza,
pero se retiro del elenco, encontró mejores ofertas”

Observamos que las redes sociales, los portales de noticias y blogs, son muy analizados, brindando información sobre qué se esta hablando y de qué forma lo están haciendo.

Nos preguntamos si en el **universo audiovisual**, resulta una fuente valiosa donde encontrar estructuras relevantes sobre el discurso mediático.



Quiénes son las personalidades más frecuentes en los informativos de TV?

Cuál es el balance de género?

A quiénes mencionan?

Qué temas abordan?

Con qué sentimiento se asocian los temas abordados según su emisor?

El proceso

1

Adquisición de datos

Transcodificación

2

Procesamiento de Video

Face recognition

Gender recognition

3

Procesamiento de Audio

Name Entity recognition

Topic Modeling

Sentiment Analysis

4

Resultados

Comparación

Visualización

5

Conclusiones

Alcances

Próximos pasos

1

Adquisición de datos

Transcodificación

C5N

El Diario

5 muestras

2hs. x muestra

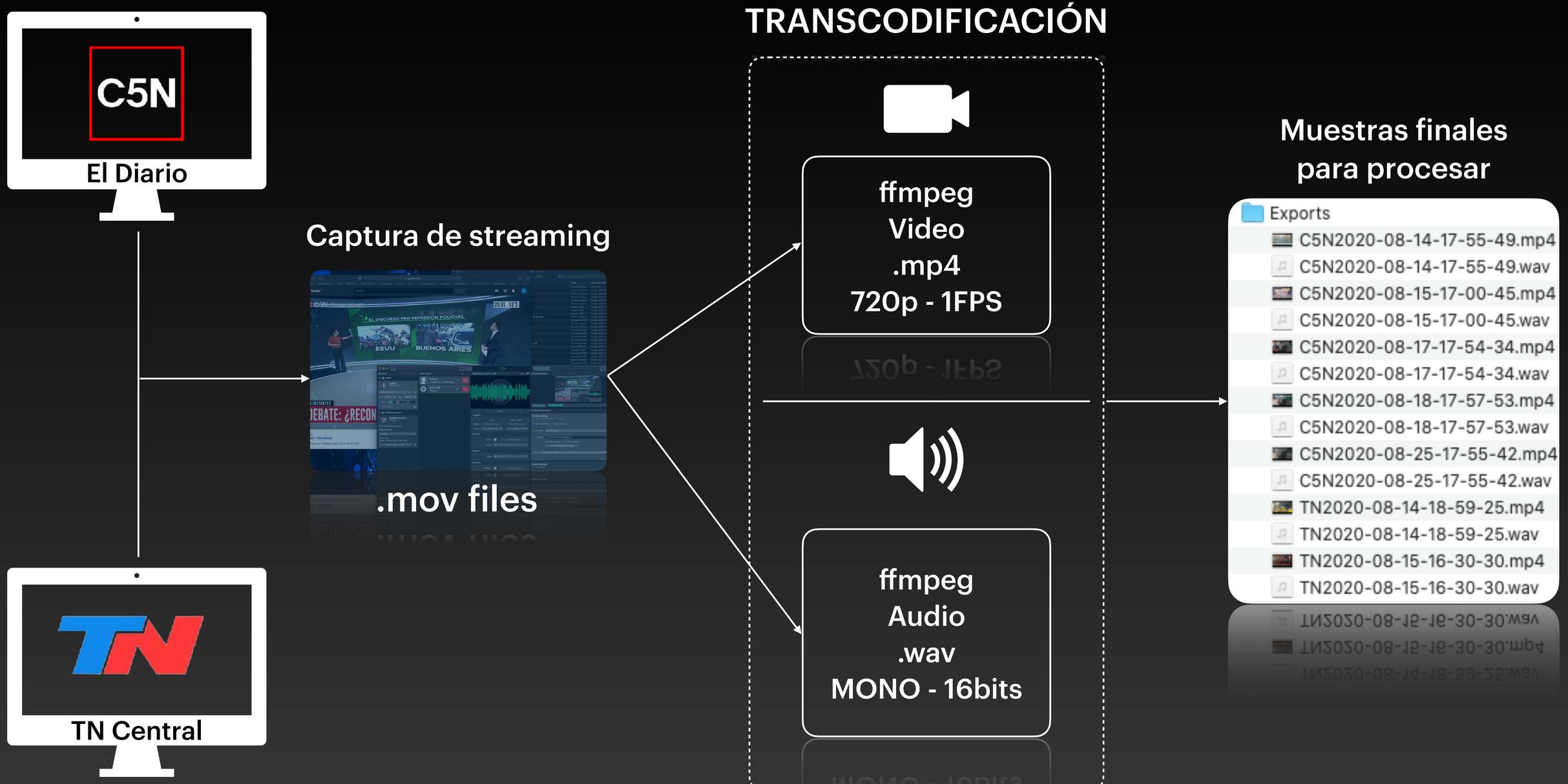
Muestreo Total
20hs.

TN

TN Central

5 muestras

2hs. x muestra



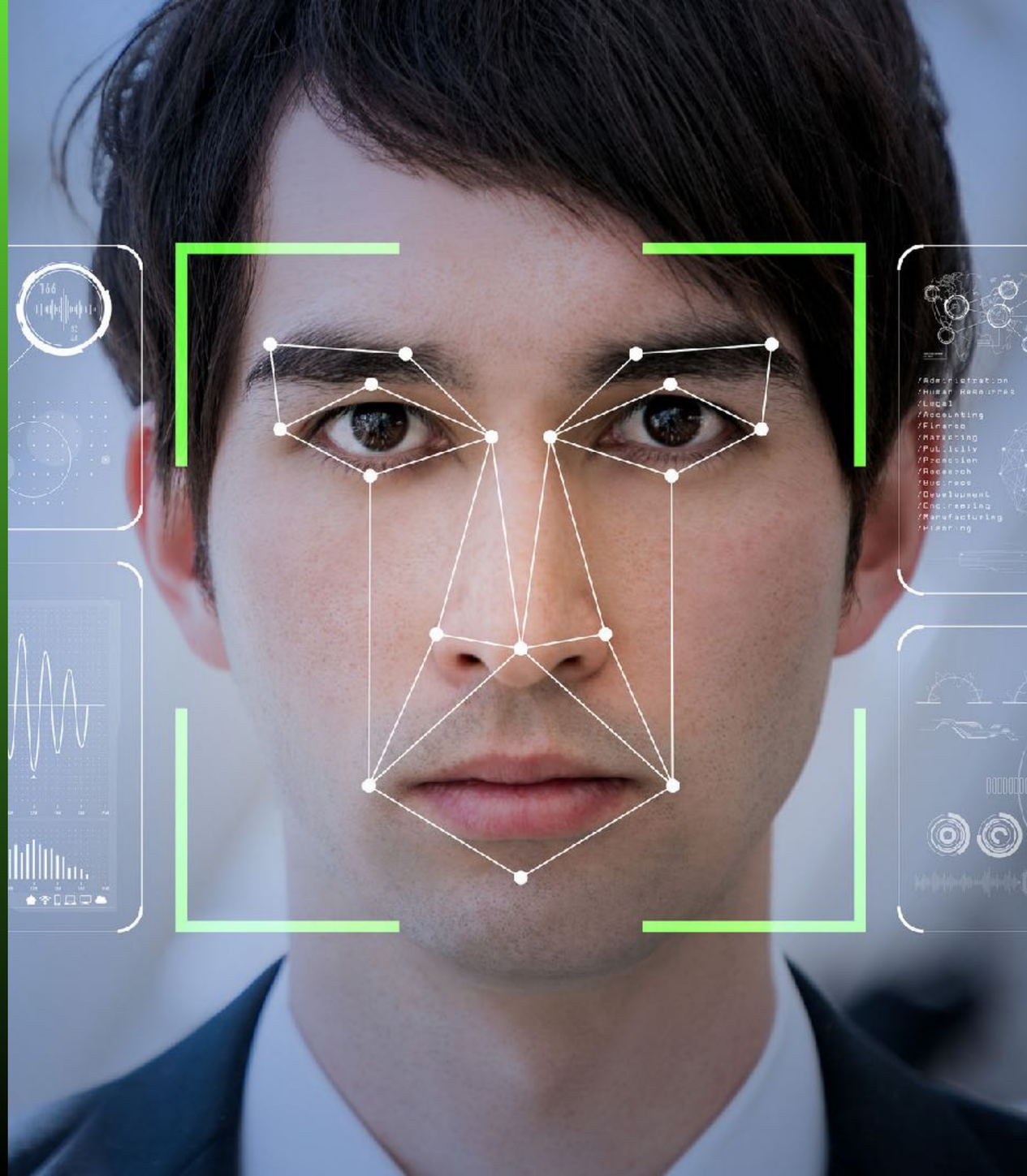
1 Adquisición de datos - Transcodificación

2

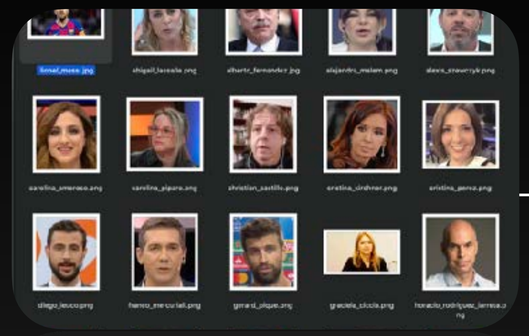
Procesamiento de Video

Face Recognition

Gender Recognition



Carga DDBB



Mp4

LECTURA ARCHIVO → Obtiene Frame → Detecta Caras → Identifica Caras → Determina Género → CSV

Caras reconocidas y géneros

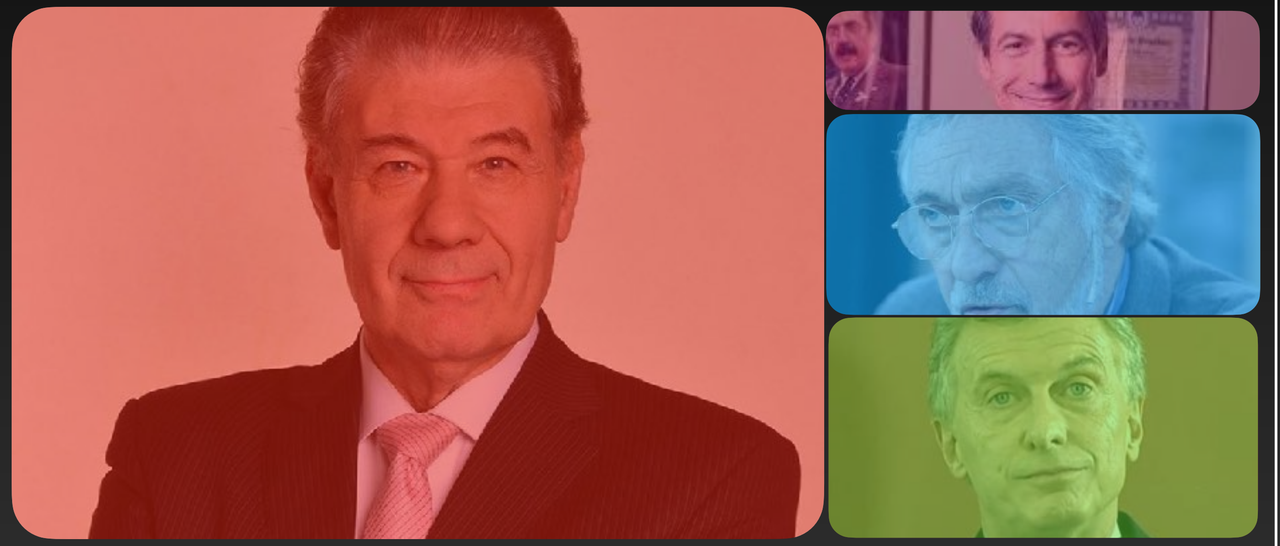
	file	personaje	genero	segundos
138	C5N2020-08-18-17-57-53	victor hugo morales	M	980
134	C5N2020-08-18-17-57-53	mauricio macri	M	215
130	C5N2020-08-18-17-57-53	luis brandoni	M	194
131	C5N2020-08-18-17-57-53	luis naidenoff	M	105
125	C5N2020-08-18-17-57-53	hernan lombardi	M	90
136	C5N2020-08-18-17-57-53	nelson castro	M	65
118	C5N2020-08-18-17-57-53	daniel gollan	M	60
124	C5N2020-08-18-17-57-53	gines gonzalez garcia	M	47

	file	personaje	genero	segundos
294	TN2020-08-18-19-00-19	luis naidenoff	M	559
302	TN2020-08-18-19-00-19	sergio berensztein	M	186
285	TN2020-08-18-19-00-19	gerardo morales	M	108
287	TN2020-08-18-19-00-19	hernan lombardi	M	100
291	TN2020-08-18-19-00-19	josep maria bartomeu	M	38
299	TN2020-08-18-19-00-19	nelson castro	M	38
278	TN2020-08-18-19-00-19	alberto fernandez	M	31
290	TN2020-08-18-19-00-19	jorge lanata	M	31

Resultados generales - 18/8/2020

Mayor tiempo de imagen en informativos centrales

C5N

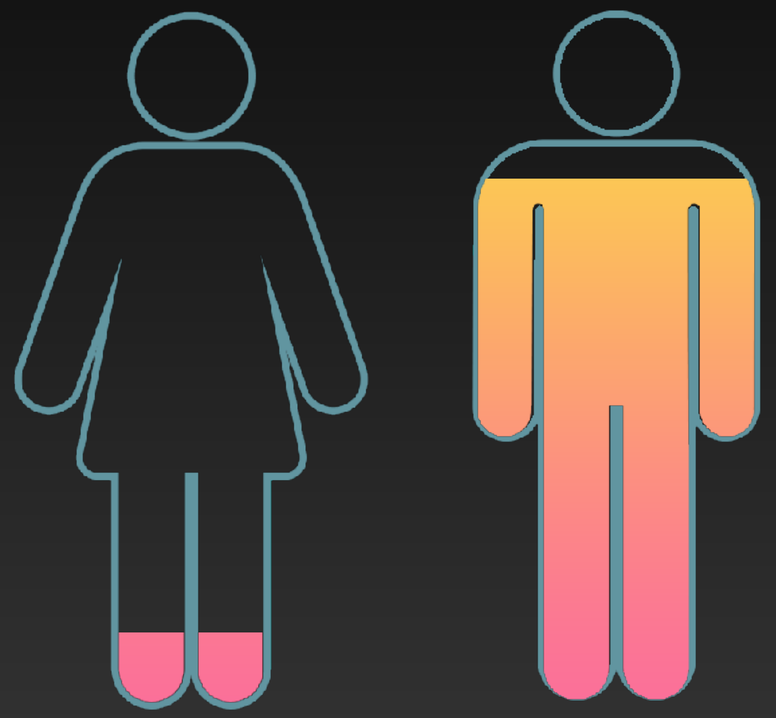


TN

Resultados generales - 18/8/2020

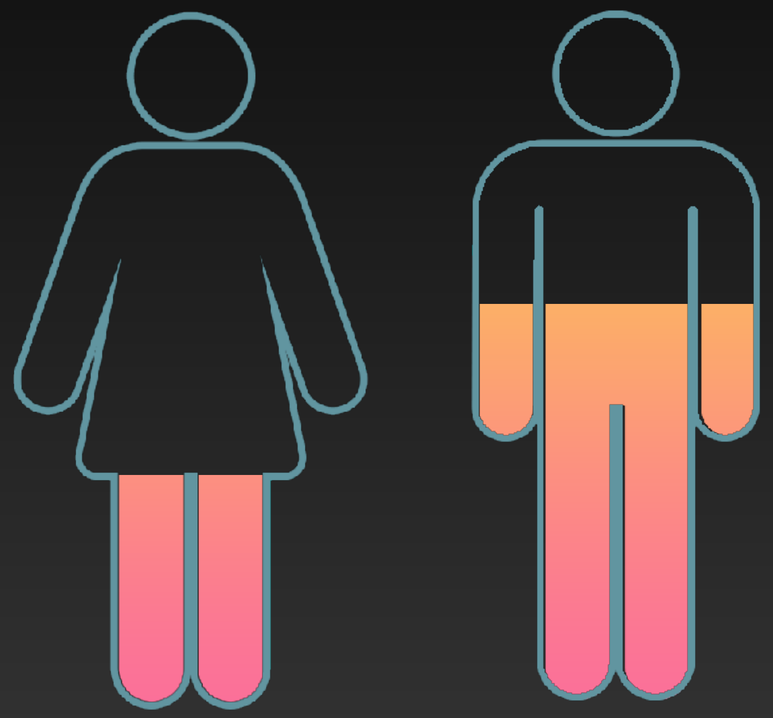
Proporción de género total reconocido

C5N



11% 89%

TN

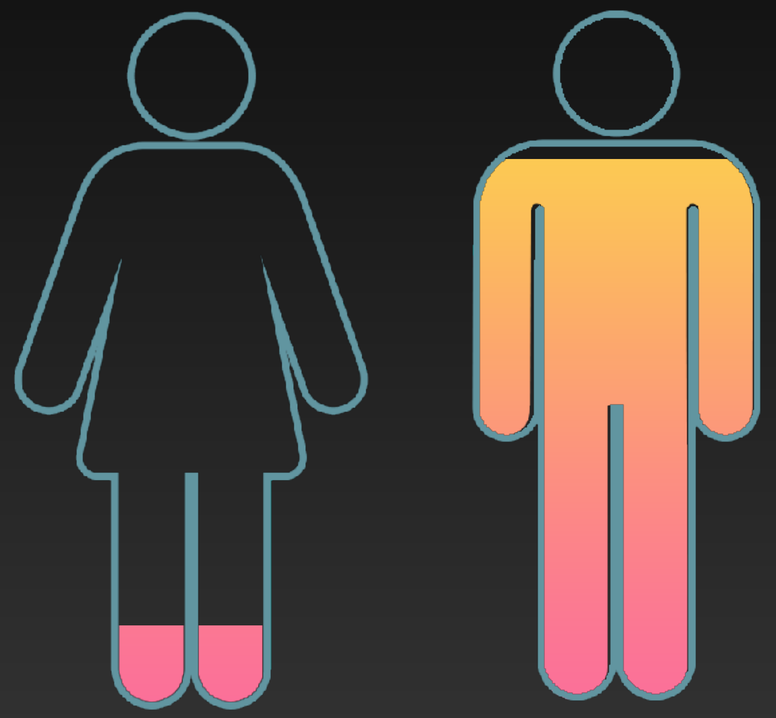


37% 63%

Resultados generales - 18/8/2020

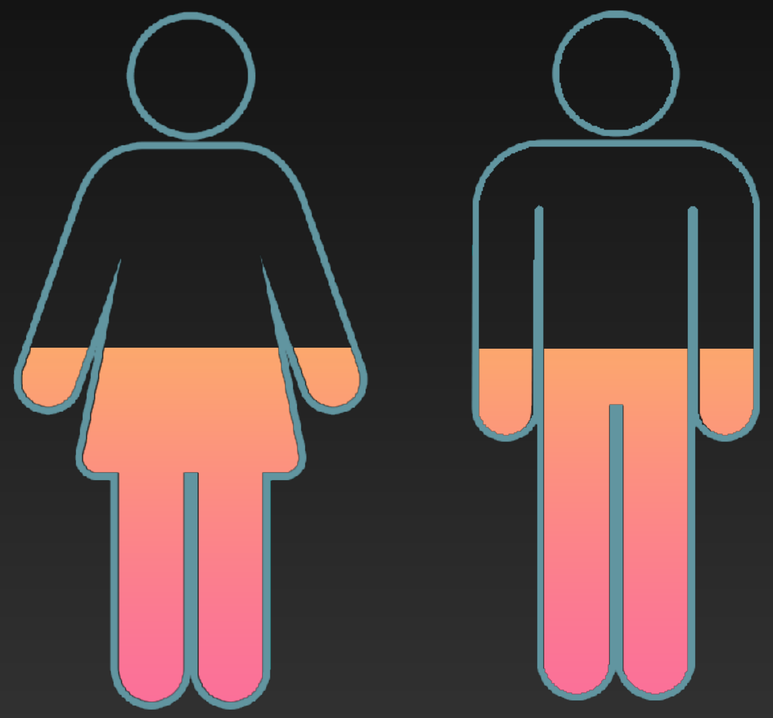
Proporción de género conductores

C5N



13% 87%

TN



50% 50%

3

Procesamiento de Audio

Name Entity Recognition

Topic Modeling

Sentiment Analysis

Extracción

LECTURA ARCHIVO → Modelo Pre-entrenado → Lee Frames → Transcribe → Obtiene resultados → CSV

"..me parece que acertó juntos por el cambio en rápidamente salir a marcar una posición en contra de la reforma judicial y alertando sobre lo que ellos consideran es el verdadero objetivo de esta reforma: la búsqueda de impunidad por parte de **cristina fernández**. Esa es la mirada de la oposición, hay que ver que pasa después y qué termina definiendo un poco la agenda pública, qué pasa en diputados. Me parece que la semana que viene va a lograr media sanción la reforma judicial en el senado, hay que ver **sergio massa** y **máximo kirchner** que impulso encuentran en la cámara de diputados, si esto avanza o si en realidad termina de comisión en comisión.."

date	text
2020-08-15 17:00:45	grande bueno aun con el inefable en estos días...
2020-08-15 17:00:45	hola hola cómo le va bienvenido buenas tardes ...
2020-08-15 17:00:45	visto y abrazo grande desde acá de todo el equ...

Name Entity Recognition

F-score: 0.93

Precision: 0.94

Recall: 0.91

	Channel	Text	pablo	gerardo morales	alberto fernández	julián	patricia	hola hola	darío baño	sánche bueno
0	C5N	hola hola muy buenas tardes bienvenidos al dia...	241	144	136	117	76	75	70	62
1	C5N	la gente circula porque los negocios se han ab...	45	0	52	0	25	0	0	0
2	C5N	cía del niño niña o totalmente inédito como ca...	0	0	118	0	0	0	0	0
3	C5N	con siri poco la hipocresía y con la mentira e...	0	0	59	0	0	0	0	0
[illegible]	TN	de los colores que se habían ido arrojó de las...	0	0	353	0	0	0	0	0

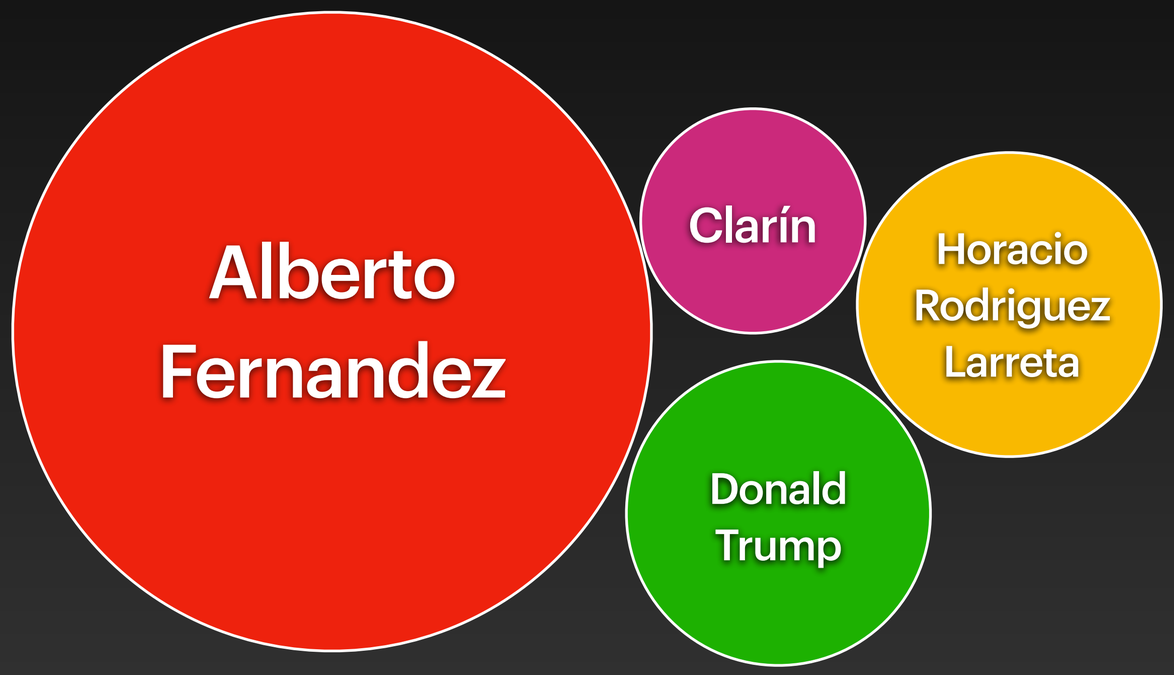
Ocurrencias

gustavo sáenz	321
macri	320
gerardo morales	293
pablo	286
mauricio macri	267
alberto	245
cristóbal lópez	195
sergio	191
hola	185
nico	176
neil	159
agustín rossi	138
fernando	136
rodrigo fuerte	136

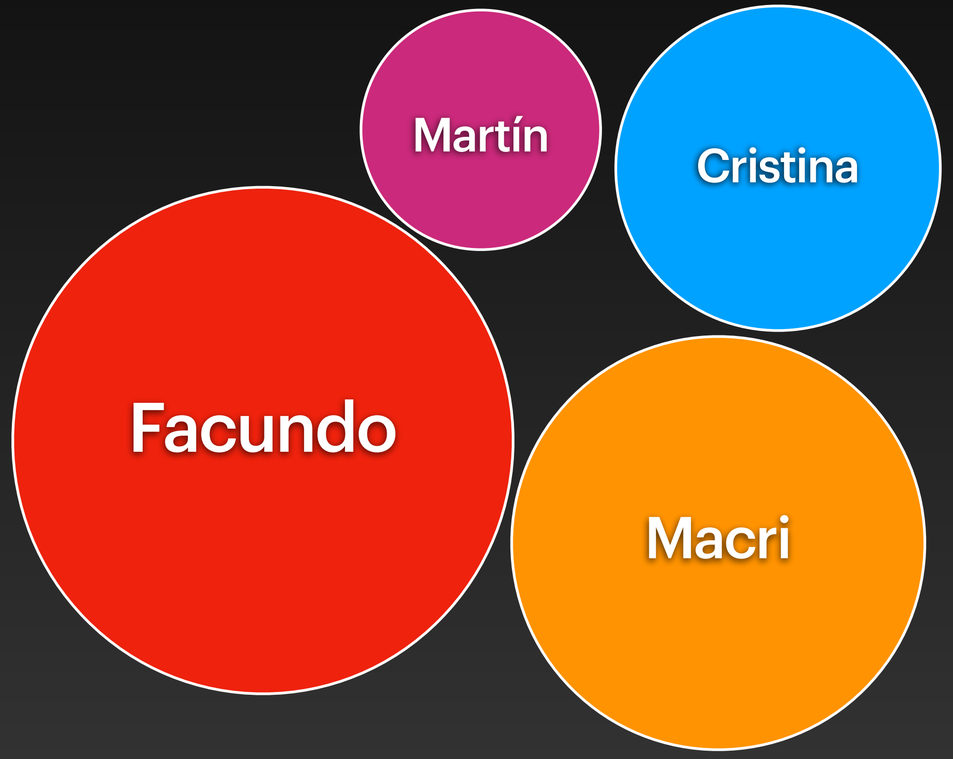
Resultados generales - 18/8/2020

Nombre mas mencionados en informativos centrales

C5N



TN



Procesamiento de audio - Name Entity Recognition - Topic Modeling - Sentiment Analysis

date	text
2020-08-15 17:00:45	grande bueno aun con el inefable en estos días...
2020-08-15 17:00:45	hola hola cómo le va bienvenido buenas tardes ...
2020-08-15 17:00:45	visto y abrazo grande desde acá de todo el equ...

Text mining

Topic

- Stopwords
- Stemming
- CountVectorizer
- TF-IDF

Tópicos

	file	topico	rank
25	C5N2020-08-18 Step 1	marcha macri cronistas atropello aduenarse mauricio suiza fifa brandon fundacion	[0.31, 0.24, 0.19, 0.16, 0.16, 0.15, 0.14, 0.14, 0.14, 0.13]
26	C5N2020-08-18 Step 2	venezuela donald trump impuestos nokia marcha pagar verguenza compara liberales	[0.26, 0.23, 0.2, 0.17, 0.17, 0.16, 0.16, 0.16, 0.15, 0.15]
27	C5N2020-08-18 Step 3	larreta marcha macri irracional temor protesta politica cercania pro policia	[0.18, 0.17, 0.15, 0.15, 0.14, 0.14, 0.13, 0.13, 0.12, 0.12]
28	C5N2020-08-18 Step 4	articulo reforma vacuna impunidad acuerdo fernandez judicial facil conmigo clinton	[0.21, 0.19, 0.16, 0.11, 0.11, 0.11, 0.11, 0.1, 0.1, 0.1]
29	C5N2020-08-18 Step 5	venezuela trump marcha macri derecho violento democracia indudablemente comportamiento ciudadanos	[0.23, 0.23, 0.15, 0.14, 0.13, 0.13, 0.11, 0.11, 0.11, 0.11]
30	C5N2020-08-18 Step 6	gabriel nieto cuerpo macri espionaje indagatoria mauricio causa guerra democratico	[0.22, 0.19, 0.17, 0.17, 0.16, 0.15, 0.14, 0.14, 0.13, 0.12]
31	C5N2020-08-18 Step 7	aburrida ejercito morales brandon democracia lorenzo jefe violaron marcha extorsiona	[0.24, 0.24, 0.21, 0.19, 0.19, 0.16, 0.16, 0.14, 0.14, 0.14]
69	TN2020-08-18 Step 1	cuerpo autopsia analisis zapatillas criminal facundo resto informacion operacion causa	[0.34, 0.28, 0.28, 0.17, 0.17, 0.14, 0.12, 0.12, 0.11, 0.1]
70	TN2020-08-18 Step 2	cuerpo facundo zapatillas policia martin mauricio hipotesis robar oposicion macri	[0.22, 0.19, 0.15, 0.15, 0.14, 0.13, 0.13, 0.13, 0.12, 0.12]
71	TN2020-08-18 Step 3	reforma judicial bloque nieto diputados senado oficial oposicion politica dictamen	[0.3, 0.22, 0.2, 0.2, 0.19, 0.18, 0.16, 0.14, 0.13, 0.12]
72	TN2020-08-18 Step 4	alexia marcha quinta senado almuerzo apurando psiquiatrico fondo reforma sesion	[0.22, 0.19, 0.17, 0.13, 0.13, 0.12, 0.12, 0.12, 0.12, 0.11]
73	TN2020-08-18 Step 5	lazaro dolar femicidios energia numero denuncia mujeres colgante omega descuento	[0.19, 0.19, 0.17, 0.15, 0.14, 0.14, 0.14, 0.13, 0.13, 0.13]
74	TN2020-08-18 Step 6	arrepentido roxana condena lazaro farina dinero casacion estero futbol santiago	[0.25, 0.21, 0.19, 0.17, 0.17, 0.16, 0.15, 0.15, 0.15, 0.13]

Resultados generales - 18/8/2020

Nube de tópicos

C5N

TN

Procesamiento de audio - Name Entity Recognition - Topic Modeling - Sentiment Analysis

Sentiment analysis

Léxico AFÍNN

Palabras negativas

	palabra	puntuacion
584	coño	-5
1894	perras	-5
1757	Niggas	-5
1893	perra	-5
2346	tragona	-5

Palabras positivas

	palabra	puntuacion
1326	Hurra	5
910	encantados	5
283	asombroso	5
1072	excepcional	5
1608	magnífico	5

	file	topico	sentimentResult
25	C5N2020-08-18 Step 1	marcha macri cronistas atropello aduenarse mauricio suiza fifa brandon fundacion	-7.1
26	C5N2020-08-18 Step 2	venezuela donald trump impuestos nokia marcha pagar verguenza compara liberales	-10.1
27	C5N2020-08-18 Step 3	larreta marcha macri irracional temor protesta politica cercania pro policia	-5.8
28	C5N2020-08-18 Step 4	articulo reforma vacuna impunidad acuerdo fernandez judicial facil conmigo clinton	3.3
29	C5N2020-08-18 Step 5	venezuela trump marcha macri derecho violento democracia indudablemente comportamiento ciudadanos	-7.9
30	C5N2020-08-18 Step 6	gabriel nieto cuerpo macri espionaje indagatoria mauricio causa guerra democratico	-8.5
31	C5N2020-08-18 Step 7	aburrida ejercito morales brandon democracia lorenzo jefe violaron marcha extorsiona	-18.7
69	TN2020-08-18 Step 1	cuerpo autopsia analisis zapatillas criminal facundo resto informacion operacion causa	-7.4
70	TN2020-08-18 Step 2	cuerpo facundo zapatillas policia martin mauricio hipotesis robar oposicion macri	-8.4
71	TN2020-08-18 Step 3	reforma judicial bloque nieto diputados senado oficial oposicion politica dictamen	-0.2
72	TN2020-08-18 Step 4	alexia marcha quinta senado almuerzo apurando psiquiatrico fondo reforma sesion	0.1
73	TN2020-08-18 Step 5	lazaro dolar femicidios energia numero denuncia mujeres colgante omega descuento	-6.2
74	TN2020-08-18 Step 6	arrepentido roxana condena lazaro farina dinero casacion estero futbol santiago	-8.6

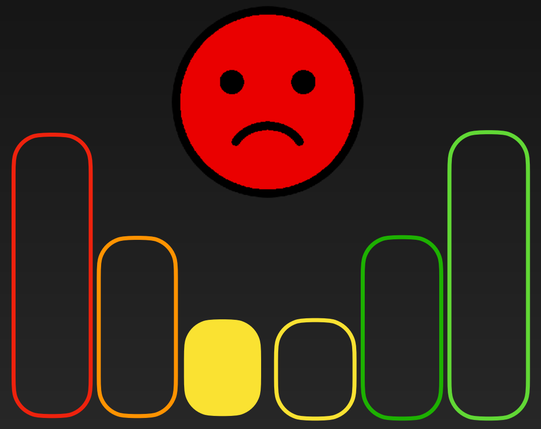
Resultados generales - 18/8/2020

Sentimiento promedio

C5N

-7.1

TN



-3.9

4

Resultados

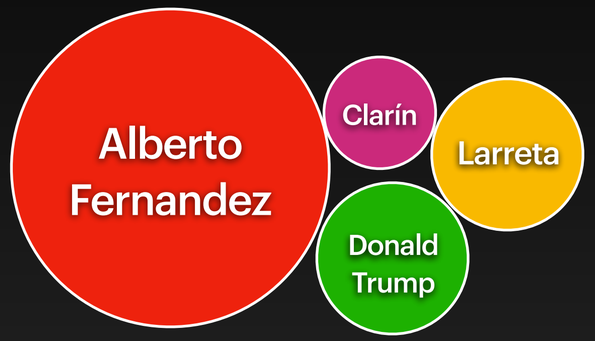
Comparación

Visualización

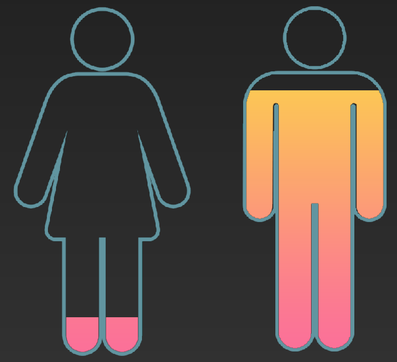
Resultados generales - 18/8/2020

Síntesis

C5N



Totales



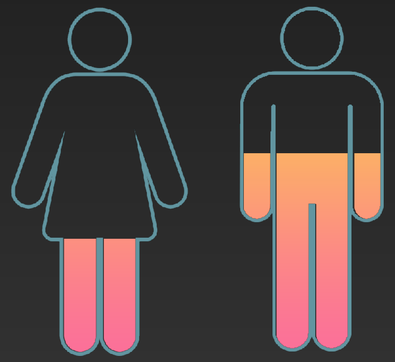
11% 89%

-7.1

TN

-3.9

Totales



37% 63%

5

Conclusiones

Alcances

Próximos pasos

Concluimos que existe una forma de **procesar el contenido audiovisual**, hallando una **fuente valiosa de información** poco utilizada en el **análisis mediático.**

El proceso aplicado a **mayor volumen de contenido**, permitiría obtener **conclusiones robustas** respecto a **quienes predominan** en la imagen, a **quienes son mencionados** con mayor frecuencia, la **balance de género** en pantalla; también encontrar en el discurso los **tópicos que dominan** la escena y con que **emocionalidad son tratados.**

Mención de Marca
Uso de Marca
Vida media de tópicos
Telecontrol
Automatización de cortes
Marketing
Uso de lenguaje
Paralelismo de tópicos
Comportamiento de audiencia
Análisis de medios

>>>Gracias()

Análisis de contenido Audiovisual

31/8/2020

Trabajo Final DataScience - Grupo 2 - DH Lima - 2020

DigitalHouse > Coding School